Between Artificial and Human Intelligence

Dr. Nuriel S. Mor
nuriel.mor@gmail.com
+972523572145

Clinical psychology, Ph.D.
Applied psychology, M.Sc.
Medical science (pharmacology & physiology), M.Sc.
Computer science, M.Sc.
Computer science, B.Sc.
Neuroscience (psychology & biology), B.Sc.
Mathematics, B.Ed.

Every so often we hear about researchers and among them Stephen Hawking who claim that artificial intelligence might be developed in such a way that could end humanity.

http://www.bbc.com/news/technology-30290540

https://en.wikipedia.org/wiki/Open_Letter_on_Artificial_Intelligence

They claim that artificial intelligence could develop human mental traits like will and self-awareness and might put human race in danger.

Already in 1957 Frank Rosenblatt the inventor of the perceptron the first learning algorithm of artificial neural networks claimed that his algorithm would create a machine that could walk, talk, see, write, reproduce itself and be conscious of its existence. This vision proved false, and perceptron is limited to linearly separable classifications without human mental qualities and self-awareness. Today with the development of deep learning, computers surpass human performance in variety of tasks such as driving, image recognition, medical diagnosis, accurate recommender systems and so on. Following this development, the claims that artificial intelligence would develop human mental features are raised again.

This view has roots in the old domain of philosophy of mind and the mind-body problem. The question is what is the nature of the relationship between body and mind. There are a physical material and measurable aspect and also a subjective private aspect of mental states. The problem is how do these two aspects interact (if at all). For this article, I will bring two main approaches which deal with this problem: ***Interactionist dualism*** and ***Materialism monism***. According to interactionist dualism mind and body are entirely two different components which interact. According to materialism monism, there is an only body (material) and all subjective mental features such as will, desires, values, self-awareness, and so on are material and derived from the complexity of the nervous system.

Therefore, the view of Stephen Hawking and others who claim that artificial intelligence could develop human mental traits such as will, values, and desires is understood in light of materialism. There is no prevention that complex artificial intelligence would develop human mental features because human mental features are material and derived from the physical complexity of the human brain and nothing more.

Here is the place to present the analysis of the great philosopher Yeshayahu Leibowitz. According to Leibowitz mind-body problem cannot be solved by humans, because the solution is beyond human perception. There is an obvious coupling between mental states and brain activity, and each subjective mental phenomenon has a biochemical expression in the brain. However, the relationship between mind and body cannot be resolved and they are completely two different components. One is material, physical, measurable in the public domain of recognition and the other is subjective, private, in the individual authority of the individual.

Leibowitz continues with his analysis and argues that science is methodically limited to explain **present** phenomena based on the **past**. For example, a stone is falling now because gravity acted

on it. On the other hand, human mind or psyche acts in the **present**, influenced by a purpose in the **future**. Leibowitz used to bring the following example: The phenomenon of a teacher that stands and talks can be explained in two ways. Explanation 1: Air coming out of the teacher's lungs vibrates her or his vocal cords, resulting in sound waves that move to the students' ears, vibrations of the students' eardrums are translated into electrical impulses that reach their auditory cortex and so on. Explanation 2: The teacher talks because he or she **wants** to say something to his or her students so they will listen. The first explanation is scientific/material/measurable, the present is explained based on the past. The second explanation is mental/psyche/psychological the present (teacher's will) is explained by the future – a purpose in the future.

Science according to Leibowitz cannot say anything about the psyche because the scientific methodology is limited to explain present by past, which is entirely different from the fundamental principle of the psyche, in which present is explained by the future. Therefore, according to this brilliant diagnosis not only that materialism does not solve the mind-body problem but also artificial intelligence could not develop human psyche traits of will. Artificial intelligence cannot develop will because it is designed by material, physical and scientific principles which are methodically limited to explain present by past, and will does not belong at all to the field of physical science and the principles of physical science.

Complex and sophisticated as it may be artificial intelligence would not **want** to end humanity because it simply would never **want**.

References

Bunge, M. (2010). The mind-body problem. In *Matter and Mind* (pp. 143-157). Springer Netherlands.

Carlson, N. R., Birkett, M. A. (2017). *Physiology of behavior*. Pearson.

Descartes, R. (2008). Descartes and the Pineal Gland. *Stanford University, November*, *5*, 2008.

LeDoux, J. E. (2002). Synaptic Self: How Our Brains Become Who We Are (Viking, New York).

Rosenblatt, F. (1957). *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory.

Schmitter, A. M. (1998). Discourse on Method and Meditations on First Philosophy. The Review of Metaphysics, 51(3), 672-674.

Yeshayahu Leibowitz (1982) Body and Mind - The Psycho-Physical Problem, Tel Aviv: Published by the Meshoderet University.